ISSN 1561-5359. Штучний інтелект, 2016, № 4

УДК 004.89: 004.93

А.В. Агарков
Інститут проблем штучного інтелекту МОН і НАН України, Україна
пр. Академіка Глушкова, 40, м. Київ, 03680

ВИДІЛЕННЯ РУХОМИХ ОБ’ЄКТІВ НА ОСНОВІ ВИКОРИСТАННЯ ЧАСОВИХ ПОСЛІДОВНОСТЕЙ

A.V. Agarkov
Institute of artificial intelligence problems of MES and NAS of Ukraine, Ukraine
40, Academician Glushkov av., Kyiv, 03680

SELECTING MOVING OBJECTS USING TIME SERIES

У роботі запропоновано рішення виділення рухомих об'єктів на зображеннях відеоряду на основі формування часових послідовностей областей руху. На кожному зображенні відеопослідовності проводиться виділення областей руху, в яких відбувається значна зміна яскравості пікселів від кадра до кадра. З множини даних областей виділяються підмножини, кожен елемент яких належить різним зображенням і які утворюють послідовності за ознакою близького розташування, розмірів і часу появи. Даним послідовностям відповідають рухомі об'єкти.

Ключові слова: виділення рухомих об'єктів, детектування об'єктів.

In this paper proposed the solution of the problem of selecting moving objects on images of a video sequence. This one based on the formation of time sequences of motion regions. In each image of the video sequence, motion areas are selected in which a significant change in pixel brightness from frame to frame occurs. Of the set of these regions, subsets are distinguished, each element of which belongs to different images and which form sequences based on the proximity, size and time of occurrence. These sequences correspond to moving objects.

Key words: selection of moving objects, detection of objects.

Вступ

Одним з найбільш ранніх способів виділення об'єктів на зображеннях кадрів відеопослідовності є метод віднімання фону [1]. Даний метод заснований на припущенні, що яскравість окремого пікселя, що належить фону залишається приблизно постійною. Її зміни обумовлені нестабільністю світлочутливої матриці відеокамери і описуються нормальним законом розподілу. Якщо яскравість пікселя відхиляється від допустимого діапазону, то це значить, що даний піксель належить вже не фону, а об'єкту переднього плану. Виділені таким способом пікселі об'єднуються в зв'язані області, які відповідають об'єктам переднього плану. На рис. 1 представлені етапи виділення об'єкта переднього плану за допомогою методу віднімання фону.

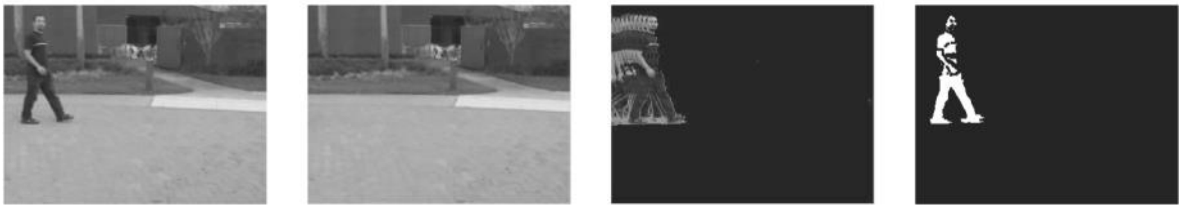


Рис. 1. Виділення об'єкта методом віднімання фону

Даний метод є дуже простим у реалізації, але досить ефективним, щоб бути використаним для вирішення різних задач комп'ютерного зору [2-4]. Даний метод можна застосовувати тільки в умовах досить стабільного фону, що не виконується для відкритої місцевості. Щоб відстежувати зміни фону, в процесі обробки відеопослідовності проводиться підстроювання допустимих значень яскравості для окремих пікселів фону. Дана процедура дозволяє зберігати ефективність виділення

об'єктів переднього плану в умовах плавної зміни освітлення і невеликих змін фону. Однак, дана процедура призводить до того, що у випадку припинення руху об'єктом переднього плану, через певний час він «зливається» з фоном і перестає виділятися.

Якщо метод віднімання фону можна віднести до методу «від противного», тобто об'єкти переднього плану виділяються за ознакою відмінності від фону, то методи, представлені нижче, призначені саме для виділення цільових об'єктів незалежно від фону.

У роботах [5-8] для детектування цільового об'єкта на зображенні використовується дескриптор, який формується для області заданого розміру. Екстрагований дескриптор аналізованої області подається на вхід навченого класифікатора. Даний класифікатор приймає рішення про те, чи містить розглянута область зображення цільового об'єкта чи ні. Дані методи використовують схему вікна, що ковзає, по всьому зображенню, що розглядається, проходять із заданим кроком по вертикалі і горизонталі, перевіряючи область заданого розміру. Для досягнення інваріантності за масштабом проводиться мультимасштабне сканування – не змінюючи розмір перевіряються області, пропорційно змінені до розміру зображення.

Найбільш відомим методом даного типу є HoG [5]. Вже згадана область розбивається на пересічні квадратні блоки, для кожного з яких формується дескриптор на основі градієнтів яскравості в кожній її точці. Даний дескриптор є гістограмою по напрямку градієнта яскравості, складеною за всіма пікселями блоку. Отримані гістограми піддаються нормалізації і складаються одна з одною в заданому порядку, утворюючи дескриптор даної області. Для прийняття рішення, чи містить розглянута область зображення цільовий об'єкт, використовується попередньо навчений SVM-класифікатор.

Розвитком методу HoG є PBM [6]. На відміну від базового методу, PBM використовує опис не тільки всього об'єкта, а й окремих його частин, що дозволяє більш точно класифікувати розглянуті області на основі моделі структури об'єкта, що описує розташування частин об'єкта на його зображенні.

Метод CENTRIST, представлений у роботі [7], практично повністю збігається з методом HoG, за винятком характеристики окремих пікселів, які використовуються для побудови гістограми. У HoG – це градієнт яскравості, в методі CENTRIST – це код, що формується на основі порівняння яскравості розглянутого пікселя і сусідніх. Даний метод працює швидше, ніж HoG.

У роботі [8] розглянута область вікна, що плавно зсувається, і описується на основі розбиття на непересічні квадратні блоки. Кожен блок відноситься до одного з попередньо визначених кластерів. Дескриптор області є гістограмою за даними кластерів по всіх блоках, що входять у дану область. Недоліком даного методу є необхідність класифікації окремих блоків і віднесення їх до одного з зумовлених кластерів.

У роботах [9,10] для виділення і відстеження об'єктів використовується сегментація оптичного потоку. Для точок кожного кадру відеопослідовності визначається оптичний потік – міжкадрове зміщення для кожної розглянутої точки. Таким чином, формуються траєкторії точок, які переміщуються від кадра до кадра. Окремим об'єктам відповідають множини точок з паралельною траєкторією. Виділення зв'язаних підмножин траєкторій дозволяє виділяти об'єкт відразу на послідовності кадрів. На рис. 2 представлені окремі кадри відеопослідовності, результати розрахунку оптичного потоку і результати сегментації множини траєкторій. Якісний розрахунок оптичного потоку вимагає достатньо великого

обсягу ресурсів, що не дозволяє використовувати метод для задач реального часу, проте підхід, заснований на сегментації множини траєкторій заслуговує на увагу.

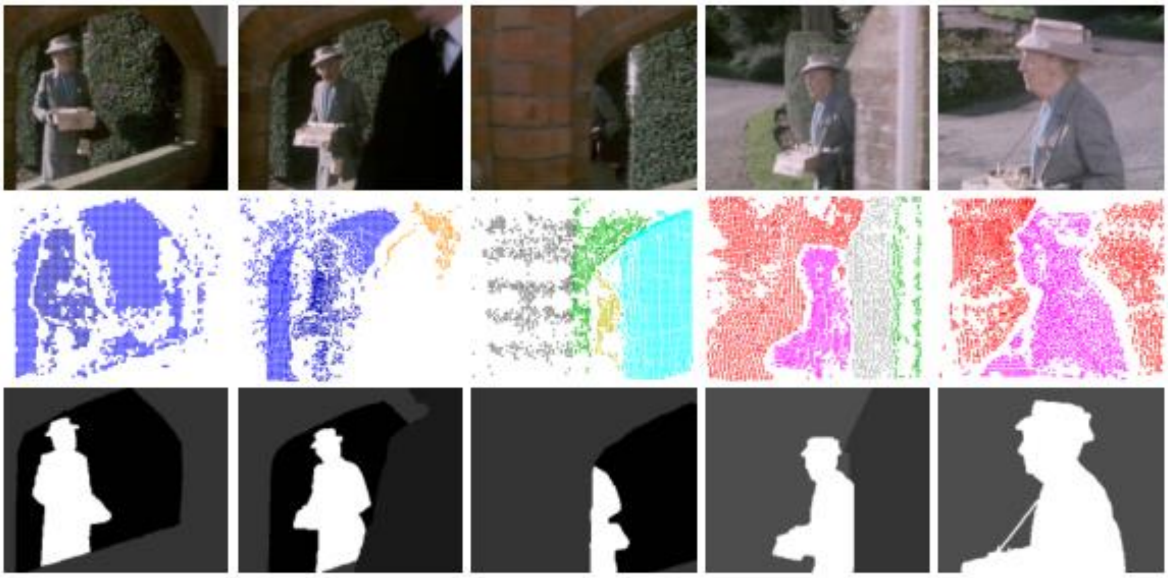


Рис. 2. Відстеження об'єктів методом сегментації оптичного потоку

У даній роботі запропоновано рішення виділення рухомих об'єктів на зображеннях відеоряду на основі формування часових послідовностей областей руху. На кожному зображенні відеопослідовності проводиться виділення областей руху, в яких відбуваються значні зміни яскравості пікселів від кадра до кадра. З множини даних областей виділяються підмножини, кожен елемент яких належить різним зображенням і які утворюють послідовності за ознакою близького розташування, розмірів і часу появи. Даним послідовностям відповідають рухомі об'єкти.

Метою даної роботи є розробка методу виділення рухомих об'єктів на основі використання обмежених обчислювальних ресурсів, для завдань первинного аналізу відеоданих.

Завдання дослідження – вивчення особливостей використання методів обробки та аналізу відеопослідовностей, що є малоресурсомісткими, для підвищення якості виконання завдання виділення рухомих об'єктів.

Виділення областей руху

Найбільш простий спосіб для виділення рухомих об'єктів – використовувати абсолютну різницю двох послідовних кадрів. При русі об'єкта яскравість окремих пікселів змінюється, що веде до великої різниці в яскравості одного пікселя на двох послідовних кадрах, ніж це може бути від шуму.

$$M_i(x,y)=\begin{cases}1, D_i(x,y)\geq T\\0, D_i(x,y)<T\end{cases},$$

де

$$D_i(x,y)=\left|I_i(x,y)-I_{i-1}(x,y)\right|,$$

$I_i(x,y)$ – i-ий кадр/зображення відеопослідовності, $M_i(x,y)$ – маска руху.

Пікселі, маски руху для яких об'єднуються в зв'язні області –

$$\Omega_k = \{p : \forall i, j; i \neq j; p_i, p_j \in \Omega_k \Rightarrow \exists L_{ij} : p_i, p_j \in L_{ij}, \forall p \in L_{ij}, M(p) = 1, p \in \Omega_k \},$$

де $p_i = (x, y)_i$ – i-ий піксель з координатами $(x, y)_i$, $\Omega_k \in \Omega$, Ω – множина зв'язних областей, L_{ij} – шлях, що з'єднує пікселі p_i і p_j.

Дані зв'язні області повинні бути повними, тобто володіти тією властивістю, що не має пікселів, додавання яких розширює дану область

$$\neg \exists p_k : (\Omega_k \cup p_k) \in \Omega.$$

Позначимо R_i прямокутник, орієнтований вздовж сторін кадра, описаний навколо області Ω_k.

Внаслідок того, що різні частини об'єкта можуть рухатися з різною швидкістю, зображення цільового об'єкта може не покриватися однією зв'язаною областю руху. Для того, щоб виділити на зображенні область, займану цільовим об'єктом, необхідно об'єднати області руху, що знаходяться поруч одна з одною. Для об'єднання областей використовуються прямокутники R_i, описані навколо них. Проводиться повний перебір всіх пар даних прямокутників. Якщо пара даних прямокутників перетинається або знаходиться близько одна до однієї, вони об'єднуються, множина прямокутників і їх пар оновлюється і процедура об'єднання триває.

$$D(R_i, R_j) < tD \Rightarrow R = (R \setminus R_i \setminus R_j) \bigcup (R_i \cup R_j),$$

де $R = \{R_i\}$ – множина прямокутників, описаних навколо областей руху.

Дана процедура повинна бути здійснена кілька разів, оскільки її одноразове виконання не дозволяє об'єднати всі області, що відносяться до одного об'єкта. Експериментально встановлено, що досить п'яти ітерацій.

Після проведення процедури об'єднання областей руху кожній виділеній області буде відповідати окремий об'єкт, що рухається (або група об'єктів).

Формування часових рядів

Вищеописана процедура дозволяє виділяти на кадрі області руху, однак не всі вони відповідають цільовим об'єктам. Наявність неправдивих відгуків пов'язана як з шумом, що виникає в матриці камери, так і динамічними змінами фону. Як правило, шумові ефекти і динамічні зміни фону носять спорадичний характер і області руху, викликані ними, нестабільні в часі, на відміну від цільових об'єктів. Області руху, ініційовані рухом цільових об'єктів, мають повторюваний від кадра до кадра (тобто в часі) характер. Дана властивість областей руху, які викликаються рухом цільових об'єктів, використовується для фільтрації помилкових цілей.

Цільовому об'єкту, що рухається, повинні відповідати приблизно однакові за розмірами області руху на послідовності кадрів. Виділення даних тимчасових послідовностей з областей руху на послідовному ряді кадрів дозволяє знайти області руху, найбільш ймовірно пов'язані з цільовим об'єктом і відфільтрувати області руху, які стосуються шуму і змін фону.

Процедура виділення даних часових послідовностей областей руху виглядає наступним чином:

$$\Theta_l = \left\{ \Omega_{l_k}^m, m = m_0 ... m_0 + N, k = 1...N : \forall 1 \le k < N, \Omega_{l_k}^{m_0+k} \approx \Omega_{l_{k+1}}^{m_0+k+1} \right\}$$

де Ω_j^i – j-та область руху на i-ому кадрі, Θ_l – l-та послідовність областей руху, $\Omega_{l_k}^{m_0+k} \approx \Omega_{l_{k+1}}^{m_0+k+1}$ – означає, що дві області руху на двох послідовних кадрах досить сильно перетинаються і мають схожі розміри. Для розрахунку схожості між областями руху Ω_1, Ω_2 використовується вираз

$$q = \frac{2S(\Omega_1 \cap \Omega_2)}{S(\Omega_1) + S(\Omega_2)},$$

де q – міра подібності між областями Ω_1, Ω_2, $S(\Omega)$ – площа області Ω.

Для ідентичних областей (у разі, якщо $\Omega_1 = \Omega_2$) $q = 1$, якщо порівнювані області не перетинаються – $q = 0$. Для дотримання умови $\Omega_1 \approx \Omega_2$ було використано виконання співвідношення $q > 0,7$.

На рис. 3 представлена блок-схема процесу формування часових послідовностей областей руху. Кожній часовій послідовності довжиною понад заданий поріг встановлюється відповідно об'єкт для відстеження.

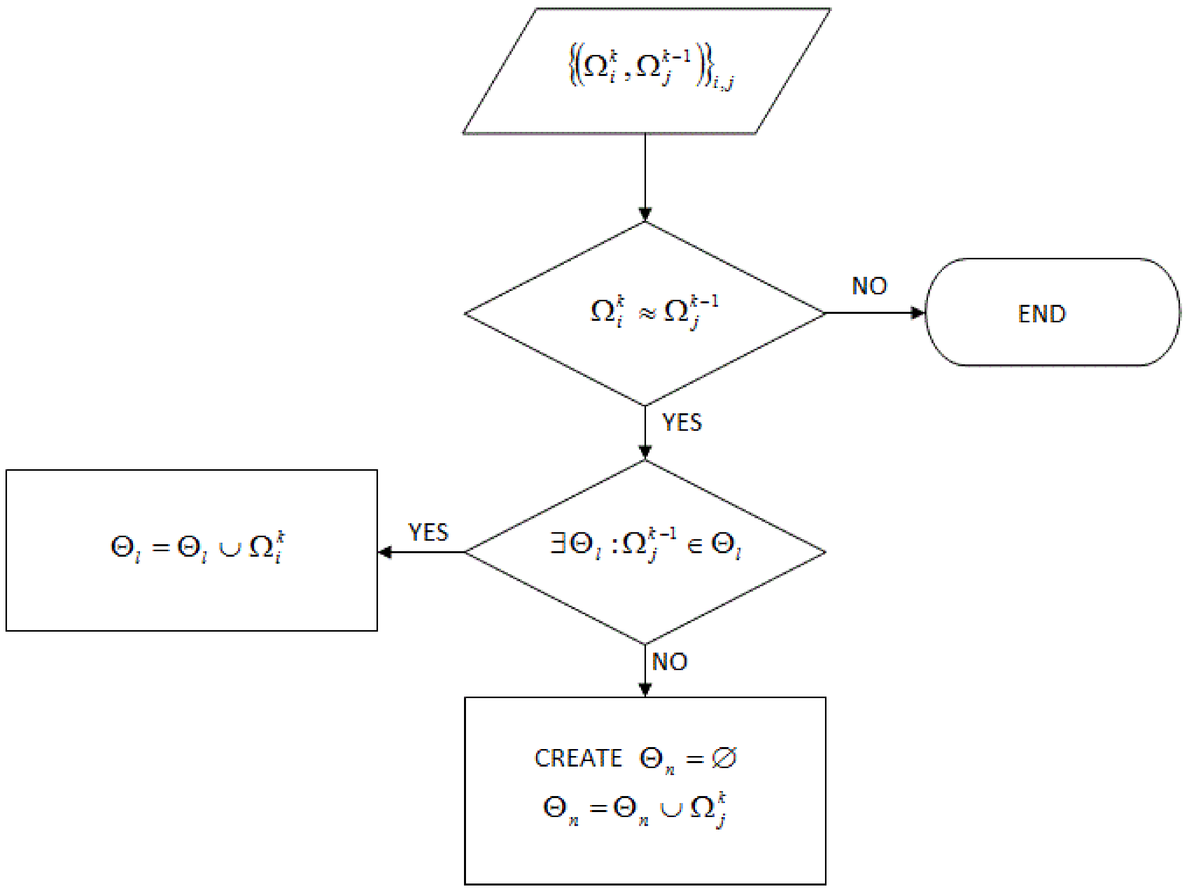
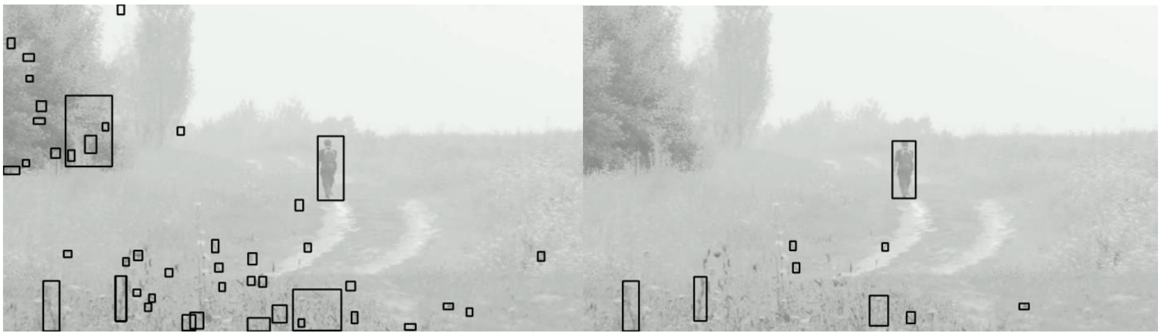


Рис. 3. Блок-схема процесу формування часових послідовностей областей руху

Експеримент

Для перевірки ефективності запропонованого методу виділення рухомих об'єктів були використані відеоролики, зняті нерухомою камерою з постійною фокусною відстанню в умовах заміського типу (дерева і трава, що коливаються, зміни освітленості). Проведені експерименти показали, що використання часових рядів для виділення рухомих об'єктів дозволило знизити кількість помилкових тривог до 5 разів. На рис. 4 представлені результати виділення рухомих об'єктів на основі виділення областей руху на поточному кадрі (рис. 4a) і з використанням часових рядів (рис. 4b). Видно, що використання запропонованого методу дозволяє значно знизити кількість помилкових тривог.



a b

Рис. 4. Результат виділення рухомих об'єктів a) на основі виділення областей руху, b) на основі часових рядів.

Висновки

Використання запропонованого методу виділення рухомих об'єктів на основі використання часових рядів дозволяє значно знизити кількість помилкових тривог без використання додаткової обробки зображень – текстурної сегментації, використання детекторів цільових об'єктів і т.п. Таким чином, отримав розвиток метод виділення рухомих об'єктів за рахунок використання часових рядів з результатів визначення руху на послідовності зображень відеопослідовності, що дозволило скоротити кількість помилкових цілей.

Література

1. Jain R. On the analysis of accumulative difference pictures from image sequences of real world scenes / Jain, R., Nagel , H. // IEEE Trans. Patt. Analy. Mach. Intell. 1979. - 1, 2. - P.206–214.
2. Wren C. Pfinder: Real-time tracking of the human body / Wren C., A Zarbayejani, A., And P Entland , A. // IEEE Trans. Patt. Analy. Mach. Intell. 19, 7,- 1997.- P.780–785.
3. Arroyo R. Expert Video-Surveillance System for Real-Time Detection of Suspicious Behaviors in Shopping Malls / R. Arroyo, J. Javier Yebes, Luis M. Bergasa, Iván G. Daza and Javier Almazán // Expert Systems with Applications. - Vol. 42. - Issue 21. - November 2015. - P. 7991-8005.
4. Dhananjaya B. Detection of Objects in Aerial Videos for Object Extraction and Tracking for UAV Applications / B. Dhananjaya, B. Rama Murthy, P. Thimmaiah // International Journal of Computer Applications (0975 – 8887). - Vol. 112 – No 12. - February 201. - P.37-42.
5. Dala N. Histograms of Oriented Gradients for Human Detection / N. Dalal and B. Triggs // Computer Vision and Pattern Recognition. - 2005. - P.886-893.
6. Pedro F. Felzenszwalb. Object Detection with Discriminatively Trained Part Based Models / Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan // IEEE Transactions on Pattern Analysis and Machine Intelligence. - Vol. 32, Issue 9. - Sept. 2010. P.1627-1645.
7. Riaz I. Human Detection By Using Centrist Features For Thermal Images / Irfan Riaz, Jingchun Piao, Hyunchul Shin // IADIS International Journal on Computer Science and Information Systems Vol. 8, No. 2, pp. 1-11 ISSN: 1646-3692.

8. Ren X. Histograms of Sparse Codes for Object Detection / X. Ren, D. Ramanan // IEEE Conference on Computer Vision and Pattern Recognition. - 2013. - P.3246-3253.
9. Brian G. Schunck. Image Flow Segmentation and Estimation by Constraint Line Clustering / Brian G. Schunck // IEEE Transactions on Pattern Analysis and Machine Intelligence. - Vol.11, Issue 10. - 2002. - P. 1010 - 1027.
10. Brox T. Object Segmentation by Long Term Analysis of Point Trajectories / T.Brox, J.Malik // 11th European conference on Computer vision. - 2010. - P. 282-295.

RESUME

A.V. Agarkov

Selecting moving objects using time series

In this paper is proposed solving the problem of selecting moving objects in video sequences. To solve this problem, it is suggested to use sequences of motion areas located on different frames, but are same by location, size, shape, time of appearance.

To allocate motion areas, it is suggested to use the absolute difference of two consecutive frames. The motion mask is formed by the method of threshold binarization with a predetermined threshold, which is chosen experimentally. The method of forest fire searches for connected areas of the motion mask. Close-lying connected regions combine and form a separate area of motion.

The sets of regions of motion of two consecutive frames are compared. If there are two areas of motion, each of which belongs to different frames, the location, size and shape of which coincide with the accuracy of the threshold, then they are considered to belong to the same time sequence. If the length of the time sequence is greater than the specified value, then the given sequence is associated with a moving object selected in the images of the video sequence. The procedure for the formation of time sequences is performed each time when the next frame of a video sequence is input, which is allowed to select moving objects in real time.

It were used video clips, taken with a fixed camera with a constant focal length in the conditions of the out of town (moving trees and grass, the variable illumination), to check the effectiveness of the proposed method for the allocation of moving objects. The experiments showed that the use of time series for the allocation of moving objects allowed to reduce the number of false alarms up to 5 times. These results show that the proposed method makes it possible to obtain a qualitative result of the selection of moving objects under conditions when the background subtraction methods based on the statistical background model give many false goals.

The use of the proposed method of allocating moving objects based on the use of time series can significantly reduce the number of false alarms without the use of additional image processing - texture segmentation, the use of detectors target objects, etc.

Надійшла до редакції 15.10.2016

